Next-Generation Conversational Interfaces

Evolution or Revolution

Giuseppe Riccardi

Adaptive Multimodal Information and Interfaces Lab
EECS Department
University of Trento, Italy
riccardi@dit.unitn.it

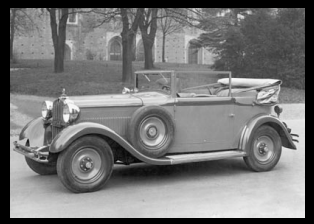
Guggenheim Museum

1959

Guggenheim Museum

1997

Speech Industry
(vs Automobile Industry)

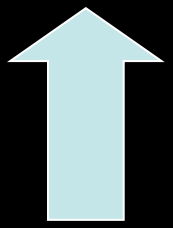


1930

1960

1990

2007



- Where is Speech technology?
  - First speech product by Treshold Technology Inc. (USA), '70s.

Conversational Interfaces

- Past, Present and Future
- Understanding Spoken Language
- Adaptive Conversational Systems
- Multimodal Interfaces
- Conclusions

Human-Machine Interaction

- The eighties
  - Top-down approach to encode (manually) knowledge of the language/world
  - Prototypes too brittle and not scalable
  - High Expectation Management (AI)
- The nineties
  - Bottom-up approach to model language/meaning/ entity relations
  - The dawn of the statistical methods
  - Lab prototypes of limited understanding machines.
  - Task Oriented
  - User interaction
  - Commercial Application!

Conversational Interfaces

(Three generations)

	1st Generation	2nd Generation	3° Generation
TTS	**INTELLIGIBLE**		
Modality	**Speech**		
Device (USER)	**Telephone**		
User Interface	**System Initiative**		
Task Type	**Command&Control**		
Machine Goal	**Automation**		

Conversational Interfaces

(Three generations)

	1st Generation	2nd Generation	3rd Generation
Vocabulary	O(1)		
Grammar	SMALL		
SLM	NO		
NLU	NO		
Dialog Models	NO		
NLG	NO		

Conversational Interfaces

(Three generations)

	1st Generation	**2nd Generation**	
TTS	INTELLIGIBLE	**NATURAL**	
Modality	Speech	**Speech**	
Device (USER)	Telephone	**Desktop** **Hand-Held**	
User Interface	System Initiative	**Mixed Initiative**	
Task Type	Command&Control	**Transactional**	
Machine Goal	Automation	**Automation**	

Conversational Interfaces

(Three generations)

	1st Generation	2nd Generation	
Vocabulary	O(1)	**O(1K)**	
Grammar	YES	**YES**	
SLM	NO	**YES**	
NLU	Grammar	**Robust**	
Dialog Models	NO	**FST/RTN**	
NLG	NO	**TEMPLATE**	

Conversational Interfaces

(Three generations)

	1st Generation	2nd Generation	3rd Generation
Vocabulary	O(1)	O(1K)	**O(100K)**
Grammar	YES	YES	**SCFG**
SLM	NO	YES	**ADAPTIVE**
NLU	FIXED/ CLOSED	Robust	**OPEN/ MULTIMODAL**
Dialog Models	NO	FST/RTN	**ADAPTIVE**
NLG	FIXED	TEMPLATE	**MODEL**

ASR Performance

(Word Error Rate)

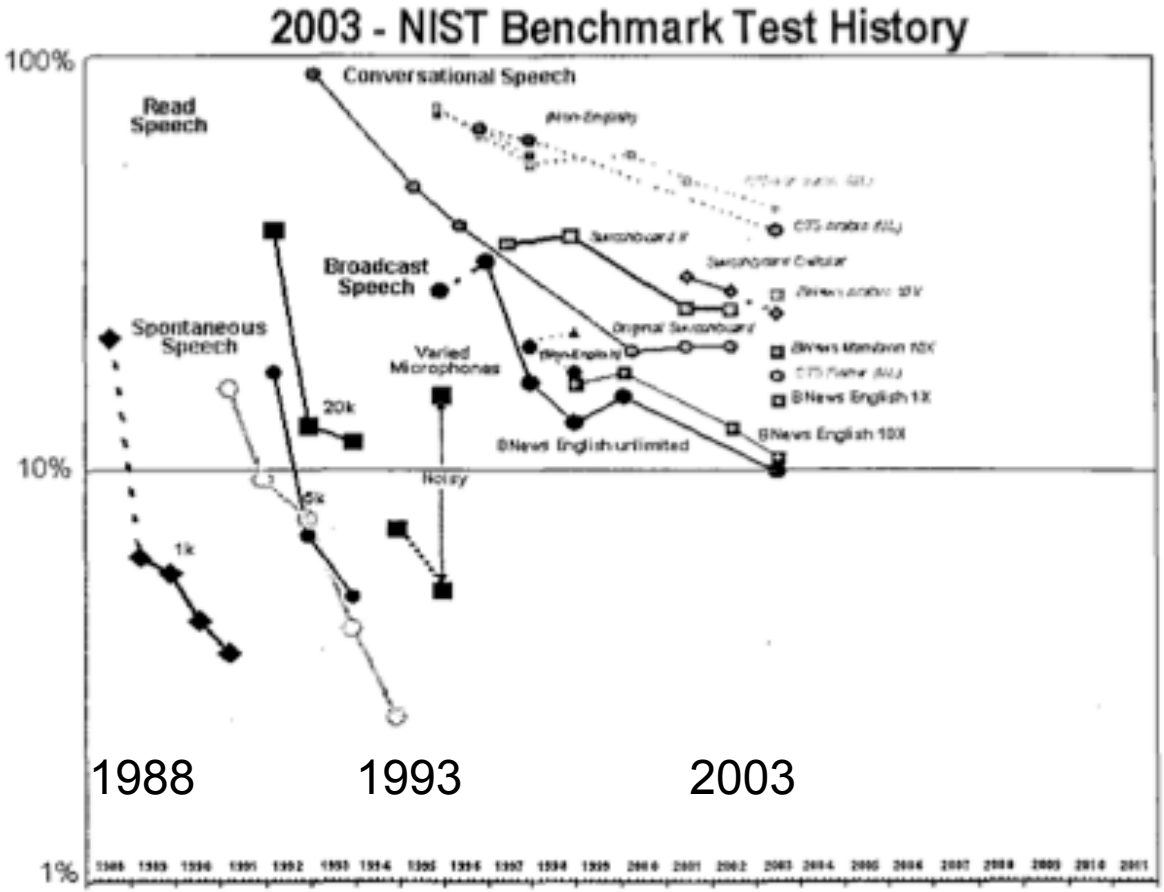


Figure 1 NIST Benchmark Test History

Conversational Interfaces

(Three generations)

	1st Generation	2nd Generation	3° Generation
Modality	Speech	Speech	**Multimodal**
Device (USER)	Telephone	Desktop Hand-Held	**Wearable/ Implantable**
Task Type	Command&Control	Transactional	**Problem Solving**
Machine Goal	Automation	Automation	**Cooperation**
TTS	INTELLIGIBLE	NATURAL	**VisualTTS** **Affective**
User Interface	System Initiative	Mixed Initiative	**Cognitive**

Types of Tasks

- Command and Control
- Informational
  - Accessing documents (web/database)
- Transactional
  - Accessing database (read/write)
- Problem Solving
  - How-To

1st Generation

2nd Generation

3rd Generation

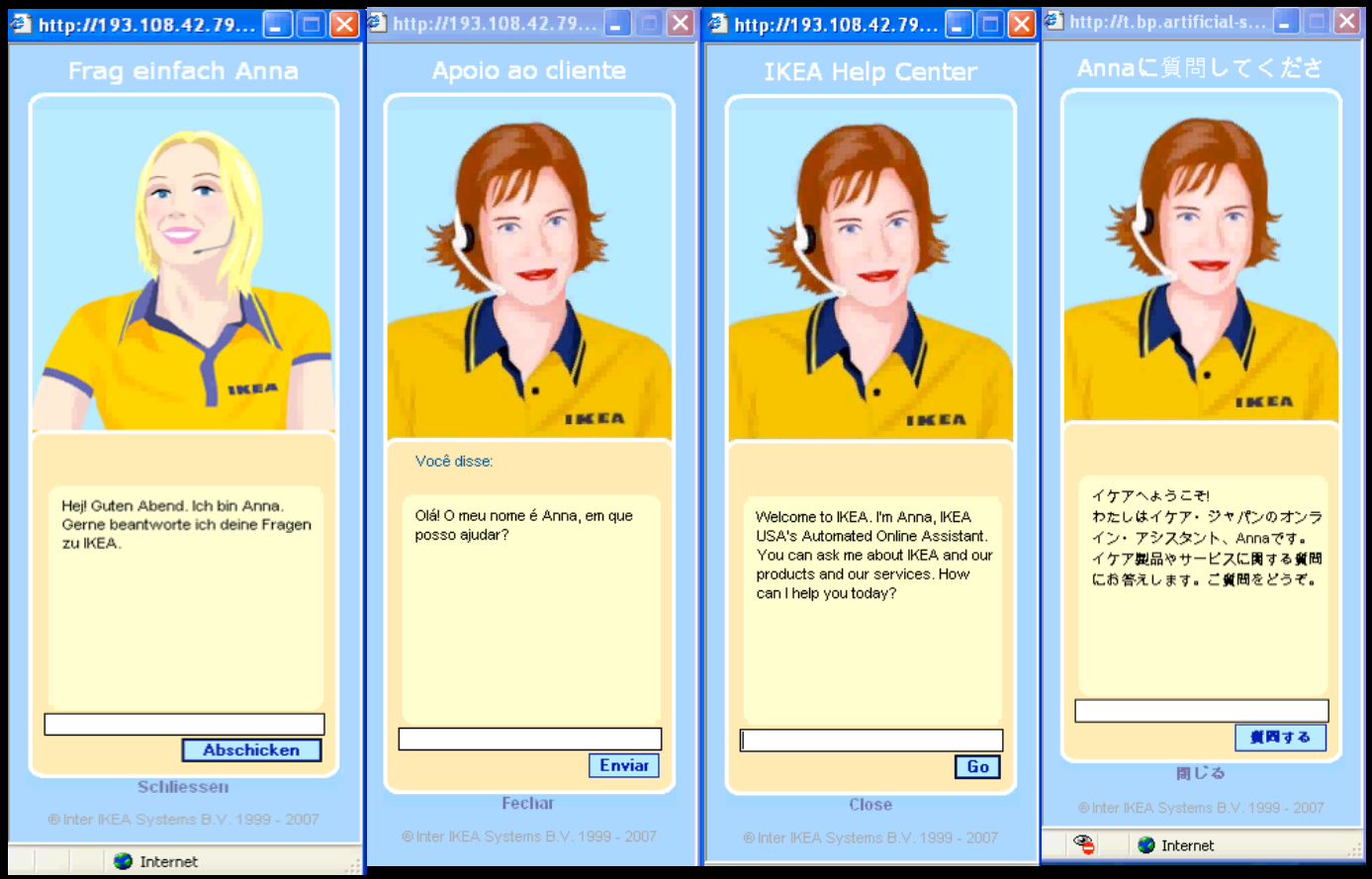
Problem Solving
(How-To)

- Problem
  - Undetermined
- Uncertainty
  - World
  - Status of the agents
- Competence
  - Cooperative task

Digital Assistants

- **Communication**
  - Speech
  - Language
  - Visual
- **User Interface**
  - Effectiveness
  - Compelling
  - Personal

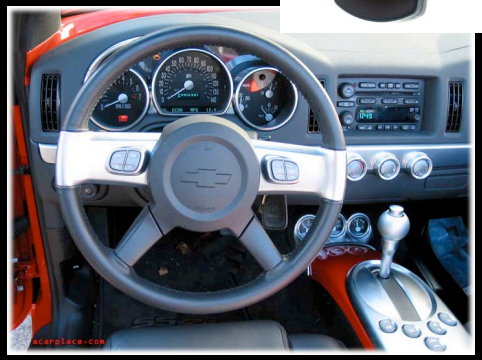
Germany Portugal USA Japan



Cognitive Interfaces

- Multiple Thread Interactions
  - Cognitive loads shift
  - Attention
  - Memory (Long tasks)
  - Emotional State (e.g angry/neutral)
- Applications
  - Problem Solving (e.g IT Help-Desk, In-Field Technical Assistance)
  - In-Car conversational interfaces

Interface Clutter

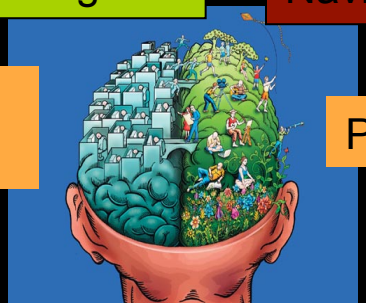


Task 1
Interacting with Tablet PC

Task 2
Driving

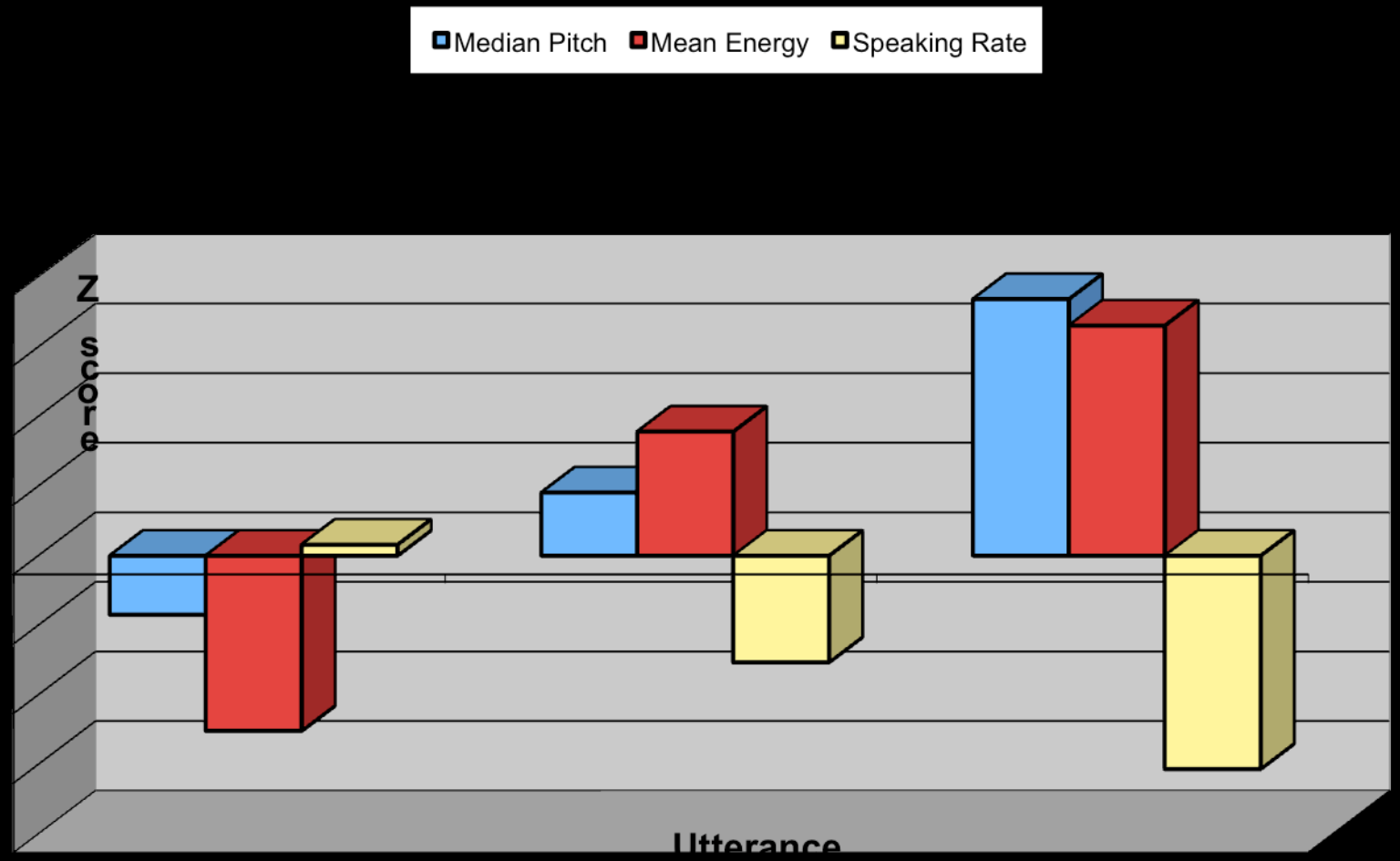
Task 3
Talking/Listening to Navigator

Task 4
Playing Music



User State Emotional Signatures

(State Transition)



Riccardi, G. and Hakkani-Tur D.,
Grounding Emotions in Human-Machine Conversational Systems ", *Lecture Notes in Computer Science, Springer-Verlag, , pp. 144 154,2005.*

Towards Third Generation Conversational Interfaces

LUNA Project

http://www.ist-luna.eu/

LUNA: Spoken Language Understanding in Multilingual Communication Systems

"The LUNA project addresses the problem of real-time understanding of spontaneous speech in the context of advanced telecom services"

Industry Partner

France Telecom
Loquendo
CSI

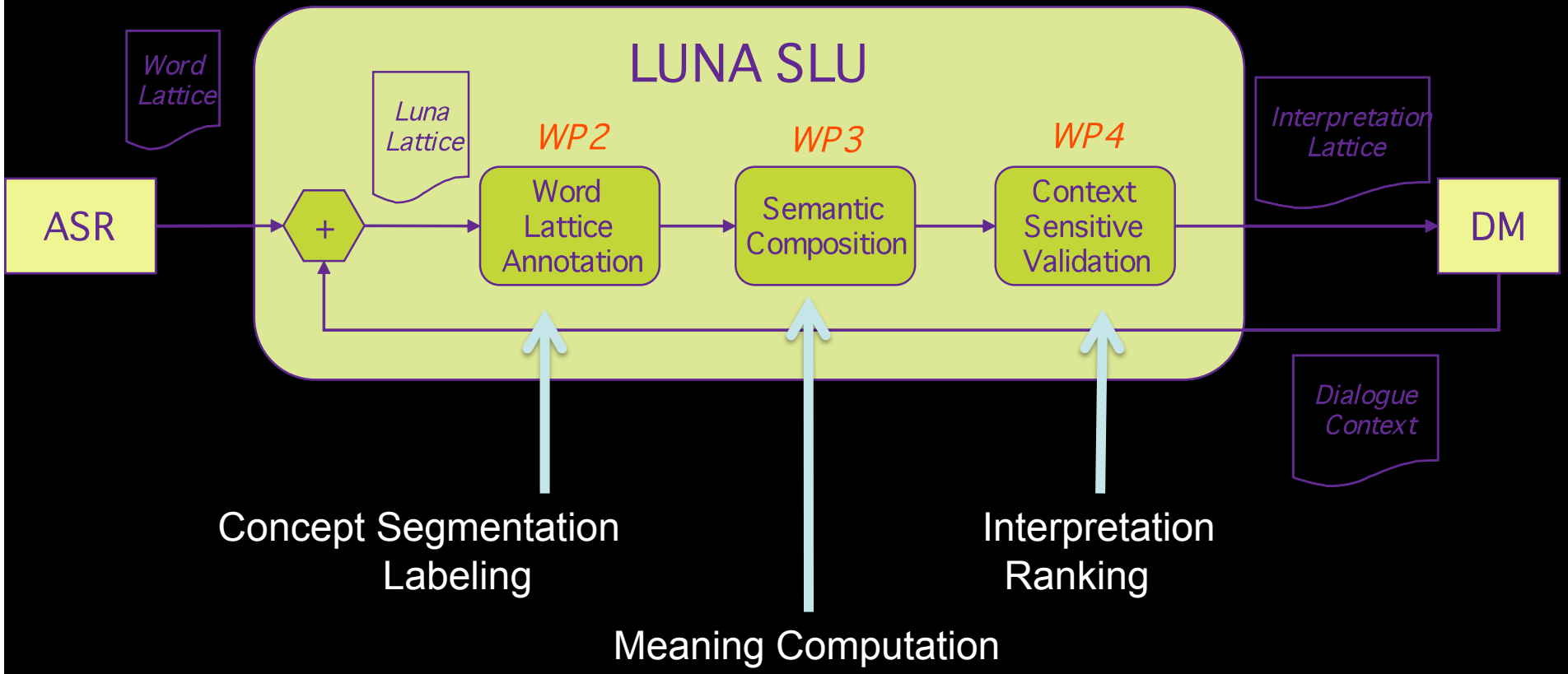
University

RWTH Achen
University of Avignon
University of Trento
Polish Academy of Science
PJIIT

LUNA Scientific Objectives

- Robust Spoken Language Understanding
  - Word-to-Concept Mapping
  - Concepts-to-Interpretations
  - Dialog Context Resolution
- Multilayer Semantic and Discourse Annotation Scheme
- Adaptive SLU
  - Active and On-line Learning
- Problem Solving Task
- From Human-Human to Human-Machine Conversations

LUNA SLU System



Human-Human Conversation

Problem Solving Task

U Hi Good Morning

O Hi, How May I Help You?

U I am Roberta Sicconi calling from Cultural Affairs at City Hall.

Personal Identification

U I had made a request for a password change yesterday

O Ok do you have the request track id?

U Uhm No I cannot find

O Ok do you have the date of the request?

U Well that was yesterday

O...ok I think I can find it..I got it

O It's for a password reset.

Problem Statement Ticket Record Retrieval

U Right. The problem is that I changed the password when I first logged in..

Problem Resolution (USER)

O You were supposed to change first time you logged in. Now let's try together to log in

O can you tell me you RVS of your computer

U Well let me see. This is a new PC to me. Where can I find it?

O Usually the tag is right next to the base of the chassy next to the power switch. It reads "inventario settore informatico".

U Inventario..Settore... Informatico. Got it 123456

Problem Resolution (PART I) OPERATOR asks help to the USER to connect to his PC

O yes that is right. Now, you see I'm writing the old login..now you type in the new login. It should be at least 6 characters...

U Ok let me write that down one moment

.....

Problem Resolution (PART II) OPERATOR and USER work together to fix the problem

Giuseppe Riccardi

Semantic and Dialog Annotation

- Domain attribute level
  Attribute-value pairs representation
  Tagset of attribute-value specified by domain ontologies
- Predicate structure
  The corpus is annotated using a FrameNet-like approach. Based on domain knowledge we define a set of frames for each domain.
- Coreference
  Different kinds of anaphoric relations (identity, bridging, set-element)
- Dialogue acts
  Initial tagset: 9 selected dialogue acts from the DAMSL scheme.

Raymond C., Riccardi G., Rodriguez K, and Wisniewska J.
The LUNA Corpus: an Annotation Scheme for a Multi-domain Multi-lingual Dialogue Corpus, DECALOG Workshop, Trento 2007
Rodriguez K, Dipper S., Götze M., Poesio M., Riccardi G., Raymond C. and Wisniewska J.
Standoff Coordination for Multi-Tool Annotation in a Dialogue Corpus,
LAW Workshop, Prague, 2007

Domain Attribute

(example)

[Operator] allora m'ha detto che [non riusciva]$_1$ ad [accedere]$_2$ [al computer]$_3$ e [le manca]$_4$ [la procedura]$_5$

trouble : [unable_to]$_1$
action : [access]$_2$
computer-hardware : [pc]$_3$
trouble : [lack_of]$_4$
computer-software : [procedure]$_5$

[User] esatto

Predicate-Argument Structure

(example)

[Operator] : allora m'ha detto che [non riusciva]$_{fe1}$ ad [accedere]$_{fe2}$ [al computer]$_{fe3}$ e le [manca]$_{fe4}$ [la procedura]$_{fe5}$

frame : access
  frame-elements : {user, hardware}
    fe id:fe1 f-element: negation
    fe id:fe2 f-element: target
    fe id:fe3 f-element: hardware
frame : need
  frame-elements : {user, requirement}
    fe id:fe4 f-element: target
    fe id:fe5 f-element: requirement

FrameNet for Speech

Model, Annotation

- LUNA dialogues with FrameNet annotation
  - Investigate to what extent FrameNet can be used for SLU (***beyond A/V representation***)
  - Scalability/Portability of Semantic Resources and Parsers
- General approach
  - Annotate FrameNet information on (partially corrected) parse trees
  - Use/augment the frames available in off-the-shelf database (e.g. Berkley)
  - Introduce new frames and/or make them more specific

Frame-based Annotation

Example

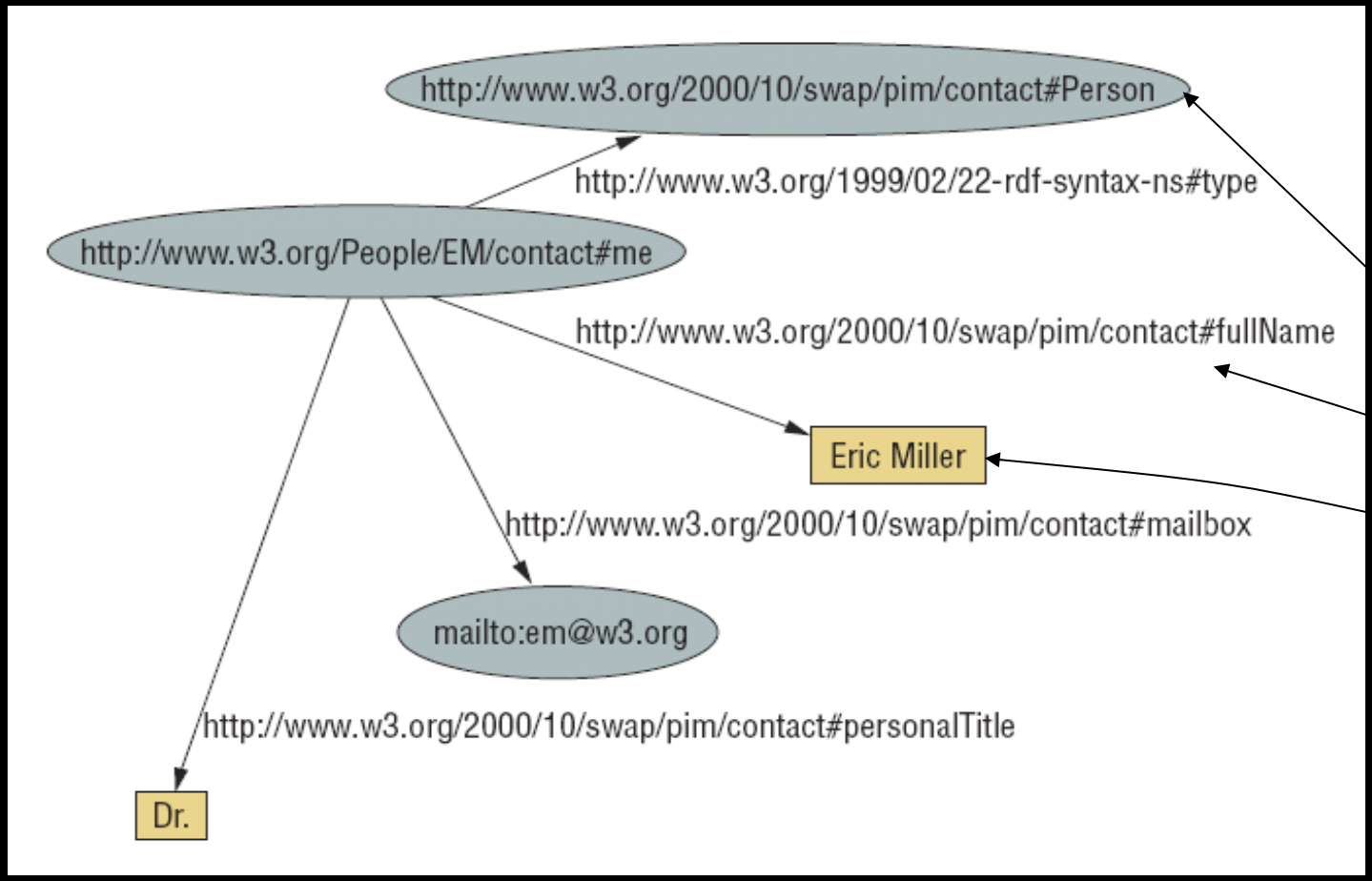
- Plain text sentence *(syntax omitted)*:
  Ralemberg said he already had a buyer for the wine.
- Target Word Selection (dictionary keyword: *buyer*)
  Ralemberg said he already had a buyer for the wine.
- Frame Disambiguation:
  Selected Frame: Commerce_Scenario
- Argument Boundary Detection:
  Ralemberg said [he] already had a [buyer] [for the wine].
- Argument Role Classification:
  *Ralemberg said [he]*SELLER *already had a [buyer]*BUYER
  *[for the wine]*GOODS.

B. Coppola, A. Moschitti and G. Riccardi
SLT workshop 2008

Open Issues

- SLU systems are critiqued for their poor domain portability, not open-domain
- How about universal representation of Semantics and/or Knowledge
- Semantic Web ?

Resource Description Framework (RDF)



Objects
Properties
Values

mailto:em@w3.org
Eric Miller
Dr

Semantic Web is not AI?

(1998)

The concept of machine-understandable documents does not imply some magical artificial intelligence which allows machines to comprehend human mumblings..... **Instead of asking machines to understand people's language, it involves asking people to make the extra effort.**

(T.B.Lee, 1998)

Web Search Queries follow this paradigm!
This is not true for Conversational Systems
Coordinate the cognitive load btw user and machine

Semantic Web Revisited

(2006)

Semantic Net of triples

object-attribute-value

Universal Resource Identifiers (URI) have global scope. *associating a URI with a resource means that anyone can link to it, refer to it or retrieve a representation of it ...***The ontologies that will furnish the Semantic web must be developed, managed and endorsed by practice communities.**

(N. Shadbolt, W. Hall, T.B.Lee, 2006)

Lessons Learned from Corpora Annotation

1% Human Error Rate in Speech Transcription

10% Human Error Rate in Sentence Annotation

Adaptive and Meaning Machines (ADAMACH)

Marie Curie Excellence Grant

Giuseppe Riccardi

Vision

- **Next Generation Conversational Machines**
  - Incremental Decoding and Interpretation of Speech
  - Domain Knowedge *vs* SLU
    - Domain/Task Ontology
    - Semantic Interpretation
  - Adaptive Dialog Models
    - Markov Decision Processes
  - Beyond words in conversational agents
    - The *persona* layer of conversational agents
    - (Social) Network of agents

Markov Decision Processes

- Modeling of Human-Machine Interaction
- MDPs vs Partially Observable MDPs
- Uncertainty in the User Input semantic interpretation
- On-line computation of best dialog strategies
- Exploration vs Exploitation

w. Sebastian Varges

Exploration vs Exploitation

- Current dialog systems do not explore, rather exploit hardwired and expensive heuristic strategies.
- Conversational Agent needs to find trade-off between exploration and exploitation
- No separation between training and testing:
  - most natural for RL and in 'real world',
  - continues to learn/adapt (learning rate)

Adaptive Learning

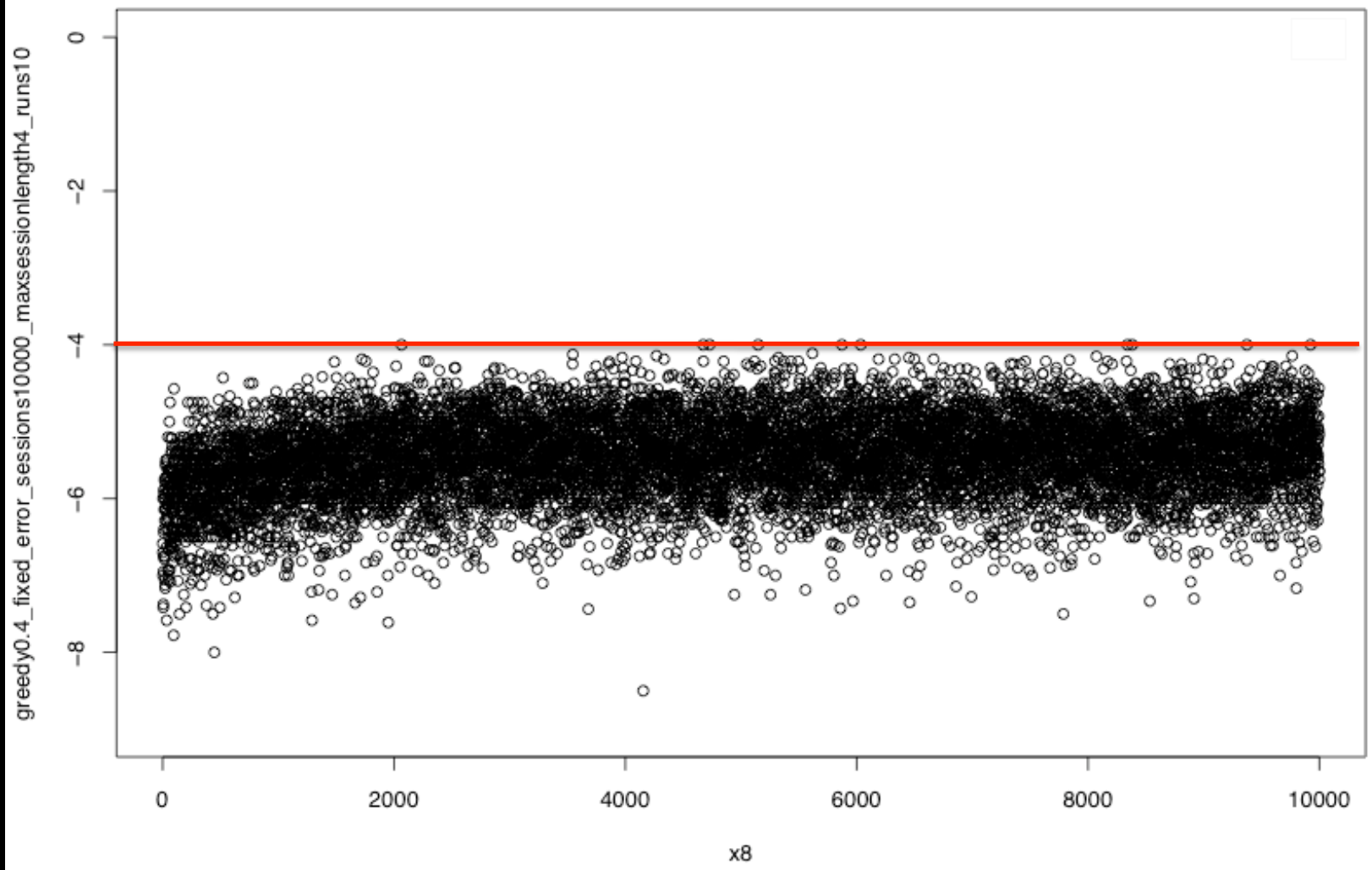
- Action selection strategy
  - Softmax (τ): actions selected according estimated probability distribution (e.g. Gibbs Distribution)

$$\frac{e^{Q_t(a)/t}}{\sum_b e^{Q_t(b)/t}}$$

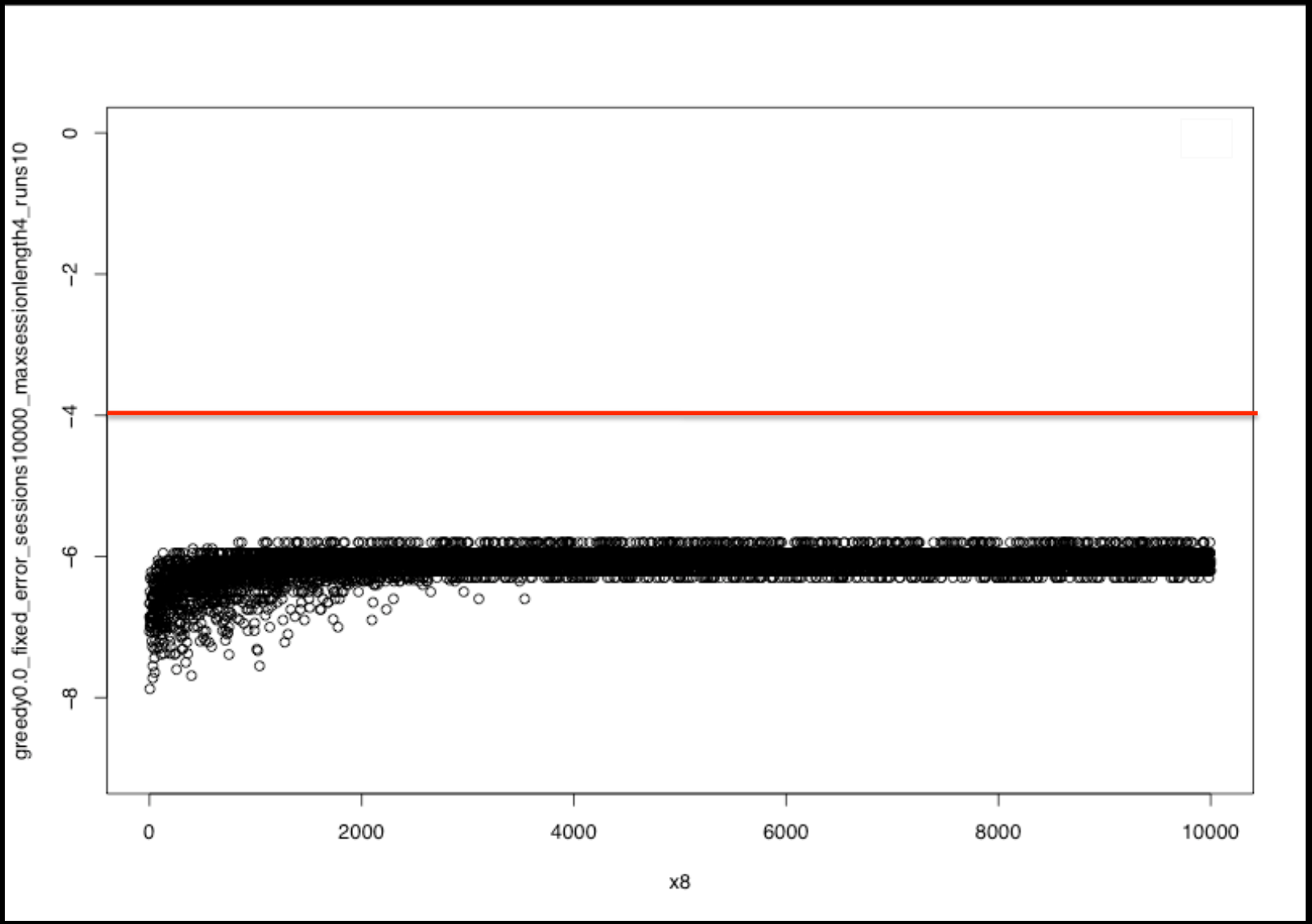
  - Greedy (ε): of random vs exploitation is selected with prob ε and exploration with prob (1-ε).
- Example
  - Adaptive Spoken Dialog System seeking to acquire two attribute slots (day and month)

40% exploration, 60% exploitation

Optimal Reward = -4



0% exploration, 100% exploitation:
Does not find optimal dialogue strategy



Vision

- **Next Generation Conversational Machines**
  - Incremental Decoding and Interpretation of Speech
  - Domain Knowedge *vs* SLU
    - Domain/Task Ontology
    - Semantic Interpretation
  - Adaptive Dialog Models
    - Markov Decision Processes
  - Beyond words in conversational agents
    - The *persona* layer of conversational agents
    - (Social) Network of agents

Personable Agents

- Role of *Personality* in communicating agents
- Current models of conversation is user agnostic at best
  - Example
- Personality modeling and generation supports
  - social layer of communication (personality matching)
    Mairesse and Walker (2007)
  - dialog strategies (e.g. content generation & selection)
  - user modeling (e.g. emotion recognition/synthesis)
- Examples (Content Generation)
  - Extrovert / Introvert
  - Introvert / Introvert

Giuseppe Riccardi

w. Adam Sporka

Social Networks

(Social) Network of Agents (1)

- Current metaphor of communication is diadic (human-machine)
- Network of agents
  - Cost
  - Social Distance
  - Trust
  - Reliability
- Humans interact with machines or a network to perform tasks or delegate them
  - Machine that interact with machines

Social Network of Agents (2)

Butler Agent

Giuseppe Riccardi

Social Network of Agents (2)

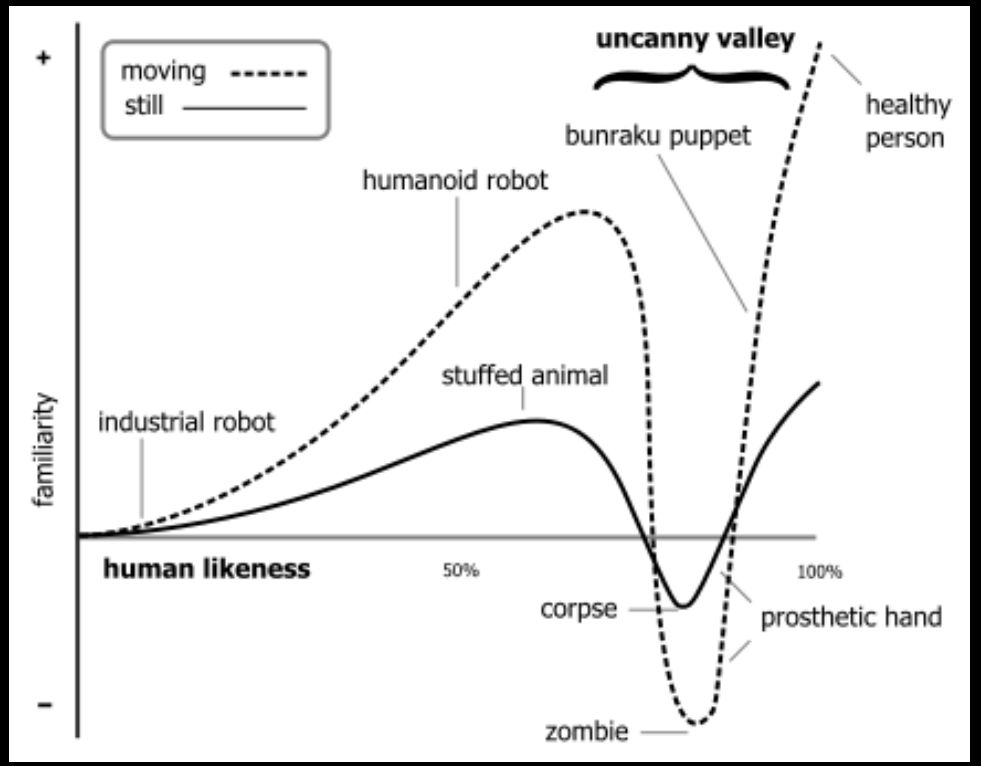
- Butler Agent
  - Delegate task such as information seeking, transactional tasks etc..
  - Dave is his name
  - YOU: "Dave check the train status of the train going from San Jose to Sacramento tonight"
  - Dave talks to other computers or human agents or both (Julie)

M-H

M-M-H

What is the limit?

- Uncanny Valley effect
  - The closer it gets to resemble human-like behavior the more likely to be rejected (non-linear)



M Mori (1970)
Ho, MacDorman, Pramono (2008)

Conversational Interfaces

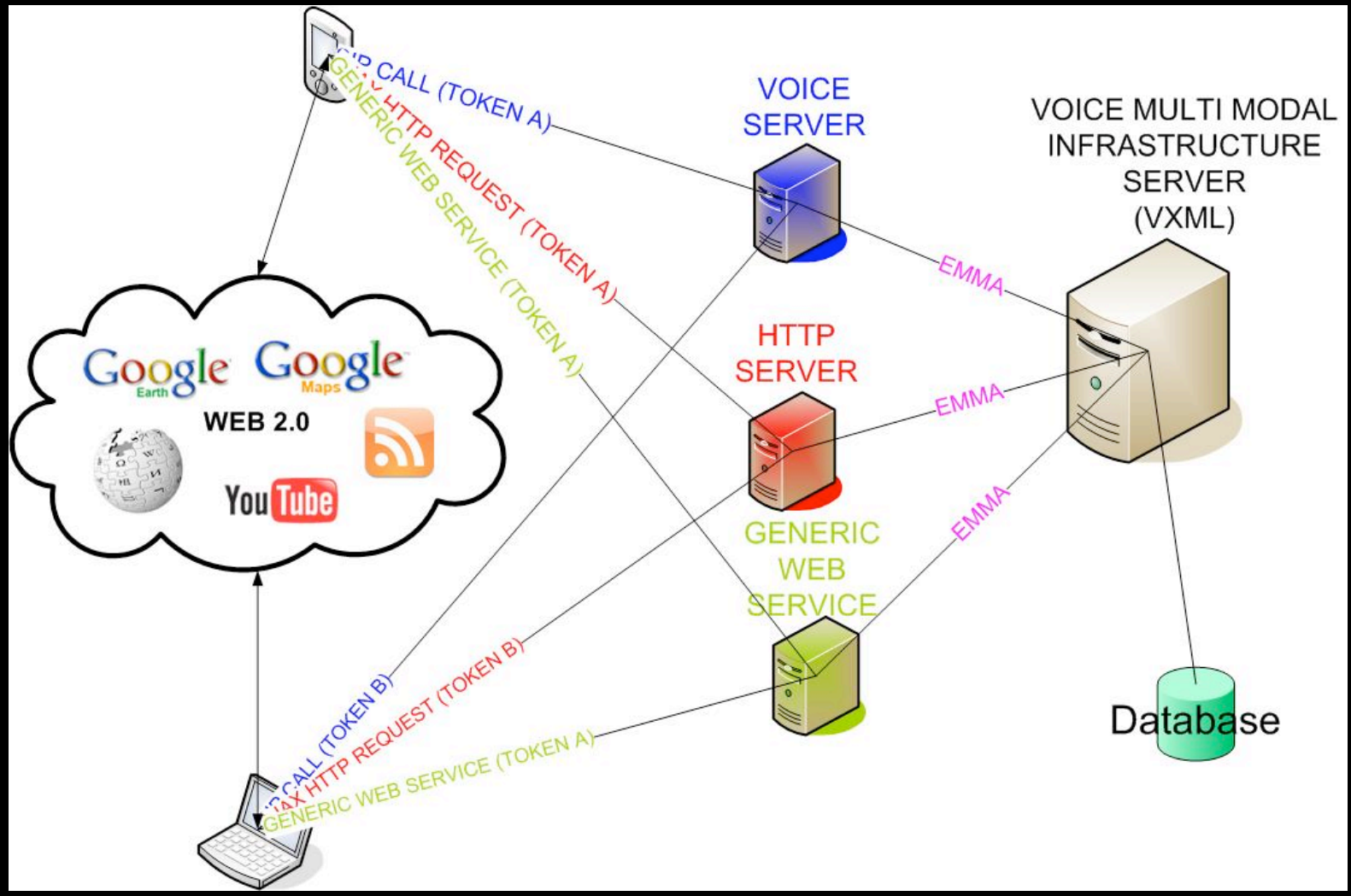
- Past, Present and Future
- Understanding Spoken Language
- Adaptive Conversational Systems
- Multimodal Interfaces
- Conclusions

Multimodal Dialog Systems

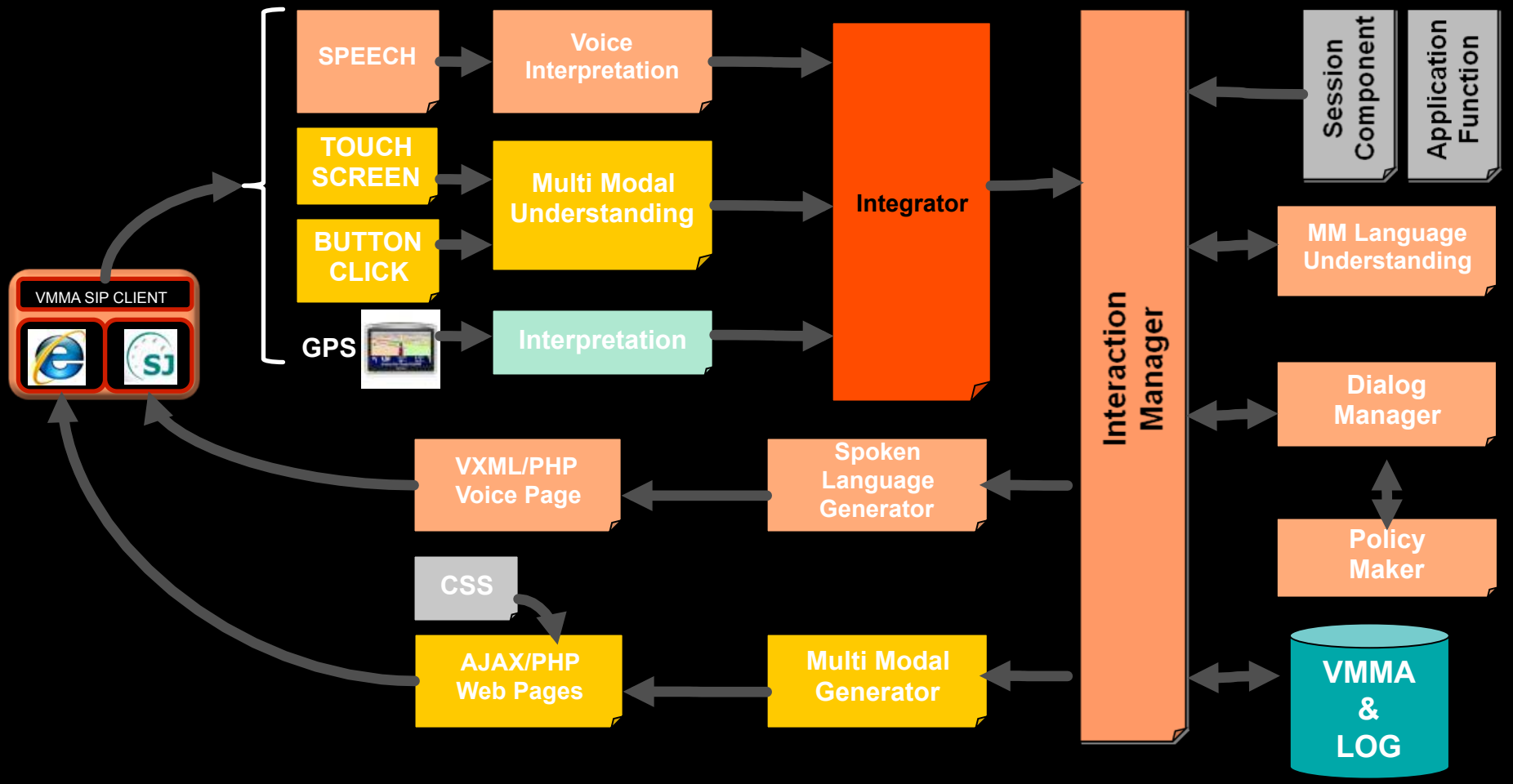
- Motivations
  - Adaptive to user/environment/task
- Architecture
- Application Framework
- Applications for mobiles users

System Architecture

casa.disi.unitn.it



Multimodal System Architecture



Giuseppe Riccardi

EMMA standard

Extensible MultiModal Annotation markup language
W3C Candidate Recommendation 11 December 2007

http://www.w3.org/TR/emma/

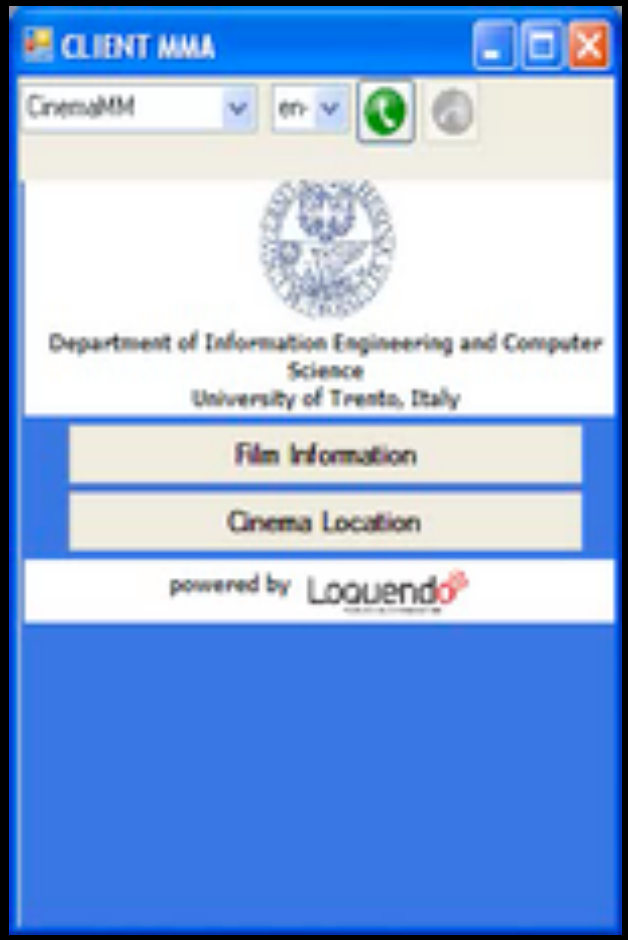
The W3C Multimodal Interaction working group aims to develop specifications to enable access to the Web using multimodal interaction.

- W3C EMMA is an XML markup language for describing the interpretation of multimodal user input
- Set of specifications for multimodal systems, and provides details of an XML markup language for containing and annotating the interpretation of user input.

Multimodal Demo

MM AutoAttendant

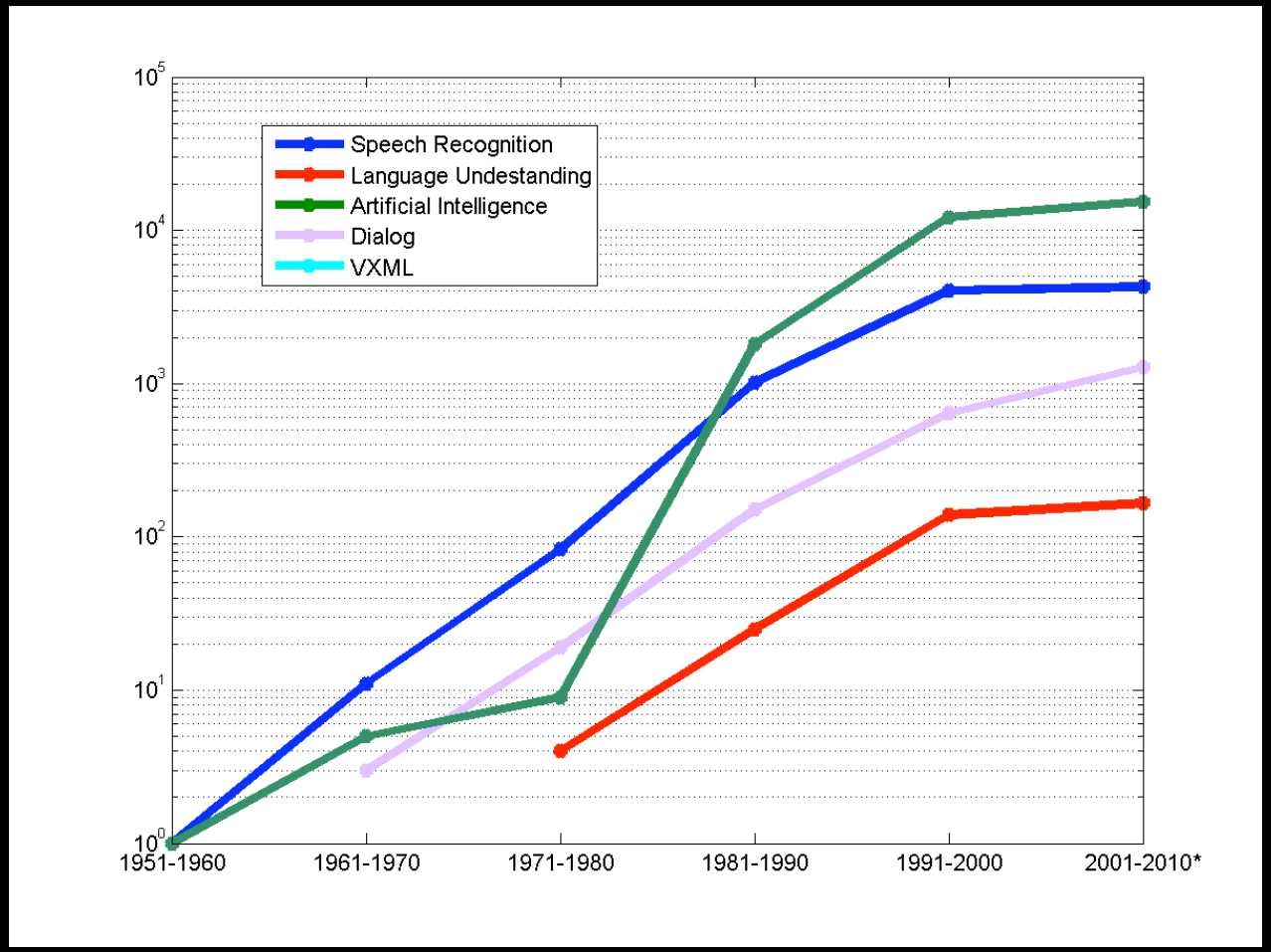
MM Movie Locator & Retrieval & Info



Giuseppe Riccardi

Research Trends

- What have researchers been working on?



(source IEEE, 2007)

Conclusion

Research Directions

- **Communicative bottlenecks**
  - Recognition vs Understanding (e.g. 10^6 ASR dictionary vs SLU 10^2 concepts)
  - Multimodal Language Understanding/Generation
  - Knowledge or Metadata (e.g. Domain ontology, un/structured database)

- **Adaptive Machines**
  - Learning Systems (active learning -> active systems)
  - Computational Models of User state as part of the "interaction equation"
  - Context-aware communication (device, environment, social)
  - User Interface Research: Personal Machines

IEEE ASRU Workshop

December 14-17, 2009

Merano, Italy

www.asru2009.org